ISSN: 2710-0588

Algorithm Improvements based on the Attention Mechanism of yolo-v5

Xuemei Wu, Wenhua Li

School of Yangtze University, Guangzhou 434000, China

Abstract

In order to further improve the accuracy of mouse species recognition, a yolo-v5 algorithm based on attention mechanism is proposed. The algorithm consists of two parts, one is the yolo-v5 backbone network, and the second part is to add attention mechanisms to the backbone network. It is used to extract global features, and then uses the attention mechanism to give different weights to the extracted features, and finally achieves the purpose of obtaining the features we need. Finally, the softmax classifier is used for classification. Experiments were validated on a homemade mouse dataset. Classification accuracy of 85% and above, respectively. Compared with other algorithms, this algorithm has a good recognition effect and robustness. Rat categories can be identified more accurately.

Keywords

Mouse Species Recognition; Attention Mechanism; yolo-v5 Algorithm; Convolutional Neural Network; Deep Learning.

1. Introduction

Image classification algorithms are commonly used in vgg16, resnet, googlenet, etc., the main function is to identify whether it is a large category of mice, effect preferences. However, if the sub-category of rats is further accurately subdivided, there are mainly different types of rats that appear at the same time.Using the yolo-v5 algorithm can improve the accuracy rate and avoid the simultaneous occurrence of different types of rats. There is no application in the field of mouse recognition, but face recognition and mouse recognition have similarities, Xiao Yani [1] et al. proposed a pedestrian re-recognition algorithm with multiple branches fusing local features. This paper mainly introduces the YOLO-V5 algorithm based on the attention mechanism.

2. Attention Mechanism

CBAM is a convolutional attention mechanism module that combines space and channel, given the intermediate feature map F=R^(C*H*W) as input, the CBAM module will infer the attention map along two independent dimensional spaces and channels in turn, and then multiply the attention map with the input feature map for adaptive feature optimization.In order to effectively extract the contour features of the detection features and obtain the main content of the detection target, the channel attention module is introduced, and the calculation method is as follows:

$$Mc(F)=sigmoid(W1(W0(Favg^c))+W1(W0(Fmax^c)) \quad (1)$$

ISSN: 2710-0588

where W0 belongs to R^(c/r*c), W1 belongs to R^(c*c/r), the two inputs share weights W0 and W1, the RELU activation function is followed by W0, Favg^c and Fmax^c represent the feature mapping generated on the space using average pooling and maximum pooling, H is the height, W is the width, C is the number of channels, and r is the reduction rate.

In order to accurately locate the position of the detection target and improve the accuracy rate of target detection, the spatial attention module is introduced to focus on key features, and its calculation method is as follows:

$$Ms(F)=sigmoid(f^{(7*7)}(Favg^c;Fmax^c)) \quad (2)$$

Where Favg^c and Fmax^c represent the average pooling characteristics and maximum pooling characteristics of the channel, f^ (7*7) represents the convolution operation of filter size 7*7[1].

3. Feature Pyramids

Feature pyramids currently play an extremely important role in object detection, pose estimation, semantic segmentation and other fields, which can greatly improve the performance of model algorithms. Feature maps of different resolutions represent information at different scales. High-resolution feature maps contain more small details, so they are called small-scale information; Low-resolution feature maps contain more content over a wide area, known as large-scale information. When targets of different sizes are sampled at different scales, the detection accuracy of small-scale targets will be relatively low. The proposed feature pyramid greatly solves this problem, it can have different resolutions at different scales, so that targets of different scales can have suitable feature representations. By fusing multi-scale information, the feature maps of all levels of feature pyramids have strong semantic information, which can improve the performance of algorithms for detecting small-scale objects [7].

4. YOLO-V5 Algorithm

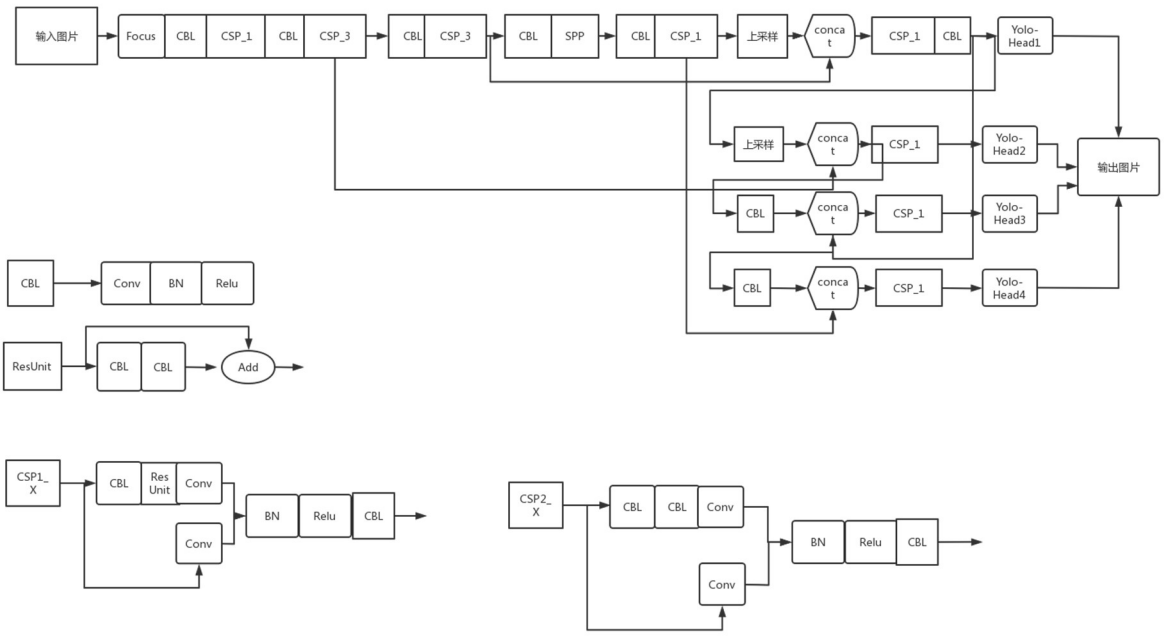

Figure 1. YOLO-V5 network structure

ISSN: 2710-0588

YOLO-V5 has the characteristics of fast speed, high flexibility and high speed, and the network structure mainly includes Darknet-53 backbone network and path aggregation network, as shown in Figure 1.

The backbone network adopts a CSP1_X structure, mainly including two branches, one branch is connected by X Crossboweck modules, the second branch is a convolutional layer, and then the two branches are spliced together, which increases the depth of the network and greatly enhances the ability to extract features.

The path aggregation network structure is a cyclic gold tower structure composed of convolution operation, upsampling operation, CSP2_X, which can make the different feature layers of the image fuse with each other for mask prediction, and obtain the final prediction box by non-maximum suppression (NMS)[2].

5. Experiment

5.1. Experimental Platform

Based on the deep learning framework built by Windows 10, Python3.9, and pytorch1.7.0, YOLO-V5 uses a 640*640 size image as input, and a feature map of equal scale size can be obtained as the detection scale. After many experiments, it is concluded that the learning rate of 0.01 can achieve local convergence faster, and the training speed is faster when the batch size is selected 16.

5.2. Experimental Dataset

This experimental dataset uses a self-made dataset. The dataset is divided into three species: Brandt's Vole, GERBILS, And Mongolian Gerbil, the images in the dataset are acquired by the camera in the rat trap, and the image resolution is 320*180, and some of the images are deleted because there are no mice, and the selected dataset has a total of 1000 pictures. The training set and the test set are randomly divided in a 9:1 ratio. In order to meet the experimental requirements, the dataset was first converted to VOC2007 format, and then the data was annotated with the help of LabelImg software, and the categories were manually labeled as Brandt's Vole, GERBILS (gerbil), Mongolian gerbil (long-clawed gerbil).

5.3. Experimental Evaluation Indicators

In order to verify the effectiveness of the extraction model, quantitative evaluation is carried out, and the main evaluation indicators are: accuracy (P), recall rate (R), average accuracy (AP), mean Average Precision (Mean Average Precision). The formula is as follows:

$$P=TP/(TP+FP)*100\% \quad (3)$$

$$R=TP/(TP+FN)*100\% \quad (4)$$

Taking the detected Brinell voles as an example, where TP indicates the number correctly identified by the detection model, FP represents the number of recognition errors or unrecognized, and FN indicates the number of Bryde voles that are incorrectly identified as large gerbils and long-clawed gerbils. P is the accuracy rate and R is the recall rate. P and R are often used to identify the quality of the model[3].

ISSN: 2710-0588

5.4. Results and Analysis

Figure 2. Contrast

It can be seen that in the case of the mouse being obscured, the model can still correctly identify the mouse.

6. Conclusion

It is of great significance to improve the accuracy and efficiency of rat species recognition, mainly aiming at the problem of partial occlusion of rats and low accuracy of species recognition, and proposes a mouse species identification method based on channel and space attention mechanism. Add the CBAM attention mechanism module on the basis of YOLO-V5, so that the neural network pays more attention to the target area containing important information and suppresses invalid information; At the same time, the data set is amplified in a data-enhanced way to improve the accuracy of small target detection.

Experimental results show that the improved YOLO-V5 model has a higher accuracy rate than the original YOLO-V5 model in the object detection task, and the detection speed meets the real-time requirements. Although the proposed method obtains performance improvement, it is still necessary to further study how to improve the detection accuracy of small objects, and the feature fusion method can be considered to improve the accuracy.

References

[1] Xing Jinchao,Pan Guangzhen. Research on Improving sign language recognition algorithm of YOLOOv5s[J/OL].Computer Engineering and Applications: 1-15[2022-06-13].http://kns.cnki.net/kcms/detail/11.2127.TP.20220510.1528.012.html.

[2] LIN Sen,LIU Meiyi,TAO Zhiyong. Using attention mechanism and improving the detection of underwater treasures of YOLv5[J].Transactions of the Chinese Society of Agricultural Engineering, 2021,37(18):307-314.

[3] WANG Lingmin,DUAN Jun,XIN Liwei. Detection method of YOLOv5 helmet wearing introduced into attention mechanism[J].Computer Engineering and Applications,2022,58(09):303-312.

[4] YANG Yongbo,LI Dong. Improved lightweight hard hat wearing detection algorithm for YOLOv5[J]. Computer Engineering and Applications,2022,58(09):201-207.

[5] Qiu Jiaohui,Pei Shaoyi,Yin Mingfeng,Qing Hongjun. Gear surface defect detection based on improved YOLOv5s[J].Modern Manufacturing Engineering, 2022(03):104-113.DOI:10.16731/j.cnki.1671-3133.2022.03.016.

ISSN: 2710-0588

[6] QIU Jiaohui,PEI Shaoyi,YIN Mingfeng,QING Hongjun. Gear surface defect detection based on improved YOLOv5s[J].Modern Manufacturing Engineering, 2022(03):104-113.DOI:10.16731/j.cnki.1671-3133.2022.03.016.

[7] Wu Xialing. Human posture estimation based on deep learning[D].Yangtze University,2022. DOI: 10.26918/d .c n k i.ghngc .2022.000106.